图书情报工作
第65卷 第19期 2021年10月

基于输入输出型SAO网络的核心技术链识别方法研究*
——以量子计算领域为例

■ 任海英　李真
北京工业大学经济与管理学院　北京 100124

摘　要：[目的/意义]基于专利文献识别产业技术领域的核心技术链，对于梳理产业核心技术架构，分析我国核心产业链的薄弱环节，确定技术攻关方向，完善技术链和产业链具有积极意义。[方法/过程]改进经典的SAO结构，提出基于输入输出型SAO网络的核心技术链识别方法。从领域专利文本中提取输入输出型SAO结构作为技术要素，根据技术要素间的输入输出关系构建技术的领域知识网络；运用加权k-Core方法获得包含主要技术关系的核心知识子网络，并通过识别强连通分量实现核心知识子网络的分解；利用主路径分析等方法识别核心知识子网络中的外部核心技术链与内部核心技术链，并参照其涉及专利对核心技术链进行解读。[结果/结论]将所提出方法应用于量子计算领域，识别出其中蕴含的外部和内部核心技术链，通过专家验证以及与相关学者研究的对比，验证了结果的准确性。该识别方法具有技术链关系连贯、自动化程度高以及灵活性强等优点。

关键词：核心技术链　SAO结构　知识网络　主路径分析　加权k-Core
分类号：G255.53
DOI：10.13266/j.issn.0252-3116.2021.19.012

1 引言

产业链是由产业上游到下游相关环节形成的链条[1]。党的十九届五中全会提出，要“推动产业链上中下游、大中小企业融通创新”，同时要“补齐产业链供应链短板，实施产业基础再造工程，加大重要产品和关键核心技术攻关力度，发展先进适用技术，推动产业链供应链多元化”。产业技术链（本文中简称“技术链”）是贯穿整个产业链各个环节的技术链条，是形成产业链的必备条件[2]。而核心技术链是一系列支撑产业活动的关键技术环节的链接，是核心产业链运行和发展的主要保障[3-4]。尤其在当前复杂的国际经济贸易环境下，打造完整的核心技术链对保障我国产业安全和推动产业链的融通创新具有重要的战略意义。因此，对技术链与核心技术链的研究越来越受到重视。

现有技术链相关研究主要包括技术链或核心技术链的概念、识别、结构、竞争态势分析以及演化机理等方面。其中，识别产业技术领域（本文中简称“技术领域”）中的核心技术链具有非常重要的实用价值，它对于分析我国核心产业链的薄弱环节、规划关键核心技术攻关方向、完善技术链和产业链都具有重要意义。

本文首先对核心技术链的相关概念和情报领域中核心技术链识别方法进行总结，分析现有识别方法的局限，然后提出一种基于SAO网络识别核心技术链的新方法，根据某一技术领域的专利文献较为全面系统地识别该领域的核心技术链，为核心技术链与核心产业链的深入分析提供一种新的方法支撑。

2 相关研究工作

2.1 技术链与核心技术链的概念

本文中的技术链是指产业技术链，主要描述一个产业中技术之间的关系。远德玉[5]较早地提出产业技术链的概念，认为它是多种生产技术的综合，是生产技术的体系化。柳琦和丁云龙[6]提出产业技术链是产业内部以及产业之间相互依存、相互衔接的技术链条。高汝熹等[7]提出技术链的两种不同描述方式：一是技

* 本文系北京市自然科学基金面上项目“基于知识网络预测和匹配的技术研发机会发现方法与评价模型”（项目编号：9192003）研究成果之一。
作者简介：任海英（ORCID：0000-0002-1197-6709），副教授，博士，E-mail：renhaiying@bjut.edu.cn；李真（ORCID：0000-0002-9469-4378），硕士研究生。
收稿日期：2021-03-30　修回日期：2021-07-14　本文起止页码：117-129　本文责任编辑：易飞

术本身的承接关系,“即一种技术的获得和使用必须以另一种技术的获得和使用为前提, 因此相关技术之间形成了一种链接关系”;二是因产品间的上下游关系而衍生出的技术链接关系,即“物化于上下游产品的各种技术依据产品间的链接关系形成了一种技术链”。黄立业等[8]提出产业技术链是将产业链的上、中、下游环节连在一起形成的技术链条。

针对核心技术链,洪勇和苏敬勤[3]提出核心技术链是包含产业活动中技术关键环节的技术链接。朱瑞博[9]认为核心技术链就是对产业核心价值链的创新和研发活动,并将核心技术链分为关键制造技术、核心元件技术和产品架构技术 3 个环节。

虽然上述技术链的定义有所不同,但都强调了技术链中各技术环节间的承接关系,而核心技术链是技术链中具有核心地位的技术链条。

2.2　核心技术链识别方法的相关研究

识别核心技术链的传统方法是专家经验法。例如,方思和李国秋[10]在参考若干学者对新能源汽车产业链和技术链的分析后,构建出新能源汽车产业链与技术链模型,并通过确定在不同产业层所使用的关键技术,识别出该产业的核心技术链。专家经验法的优点是具有较强的专业性和参考价值。然而,它的缺点是自动化程度较差,且往往受专家自身知识领域限制,具有一定的片面性和主观性。即使有多个专家,也可能因核心技术链表述的详细程度(粒度)或范围不一致而导致核心技术链的识别出现部分重叠、遗失甚至矛盾等问题。因此,近年来学者们开始采用专利分析等客观定量方法或者将定性与定量方法相结合来识别核心技术链。

2.2.1　定性与定量相结合的核心技术链识别

定性与定量相结合的核心技术链识别通常先由专家帮助找到产业的主要技术环节,然后通过对专利或其他科技情报的统计和计量对每个环节进行深入分析,找到其中的核心技术。方思和李国秋[10]通过对无人驾驶汽车产业的专利进行统计与分析,识别技术链各环节的龙头企业以及核心技术和技术分支。吴菲菲等[11]从产业技术链的角度对我国动力电池产业进行系统分析,通过产业技术链核心技术分析,剖析其中技术薄弱环节。袁晓东和鲍业文[12]通过分析中兴公司上下游合作伙伴获得以中兴公司为中心的电子及通信设备制造业技术链,然后分析各环节 IPC 共现网络中的核心专利 IPC 分类号,得到技术链各环节的核心技术主题。

2.2.2　基于专利引用的核心技术链识别

基于专利引用的核心专利链识别方法根据专利间的引用关系构建专利引文网络,通过分析引文网络中重要路径识别出其中包含核心技术传播路径的核心专利链。Y. M. Wei 等[13]构建了页岩气技术领域专利引文网络,利用 FCFP 主路径分析法识别该领域的关键技术发展路径,获得由关键技术构建的核心专利链条。赵蓉英等[14]认为专利引用关系能够体现技术的传承与发展,因此他们通过构建专利间的多级引证网络反映技术的聚合扩散情况,并以被引频次作为专利重要性的衡量,得到重要技术发展路径链条。戚筠等[15]构建了石墨烯领域的专利引文网络,将多种主路径分析算法应用于专利引文网络,识别出核心专利链条。

2.2.3　基于 SAO 的技术树或技术路线图识别

近年来,SAO 结构被广泛用于分析专利文本中技术概念间的语义关系。经典的 SAO 结构包含主体(S)、行为(A)和客体(O)。一些学者从专利文本中提取 SAO 结构,根据 SAO 结构中技术概念间的关系构建技术树、技术形态矩阵或技术路线图,这些研究为核心技术链的识别提供了新的分析工具。S. Choi 等[16]计算 SAO 结构间的相似度,通过聚类将其分为产品、技术、功能和技术属性 4 种类别,并将 AO 分为包含、效应和属性 3 类,以此构建产品、技术和功能相互关系的技术树。翟东升等[17]在 SAO 结构的基础上加入“宾补”成分,并根据 SAO 的语义特征将其标注为产品、功能、技术属性、科学效应、功效几类技术特征信息,从而构建技术领域技术树。郭俊芳等[18]通过关键概念的词性区分技术领域关键概念的技术类别,根据关键概念间的语义关系和层级关系构建 SAO 结构链,进而构建技术领域形态结构。C. Yang 等[19]将 SAO 结构中的主体(S)及客体(O)视为网络节点,综合节点间动作关系(A)的数量及其重要性程度作为节点间的边权,进而构建 SAO 网络,用于分析技术发展趋势。赖朝安等[20]根据 SAO 结构的语义特征将其分为主题、方案、结构、效果 4 类,并考虑 SAO 结构之间的技术层次,构建 SAO 结构链形式的技术树,并将其根据相似度链接构建专利网络。李晓曼等[21]构建流程化的技术要素识别框架,基于技术要素的位置特征、句法特征、动词特征、语义特征等将技术要素分为材料、产品、方法、功效和用途五类,实现对技术领域结构的梳理。

2.3　现有方法评述

上述核心技术链的识别方法实际上识别的是不同形式的核心技术链,如图 1 所示:

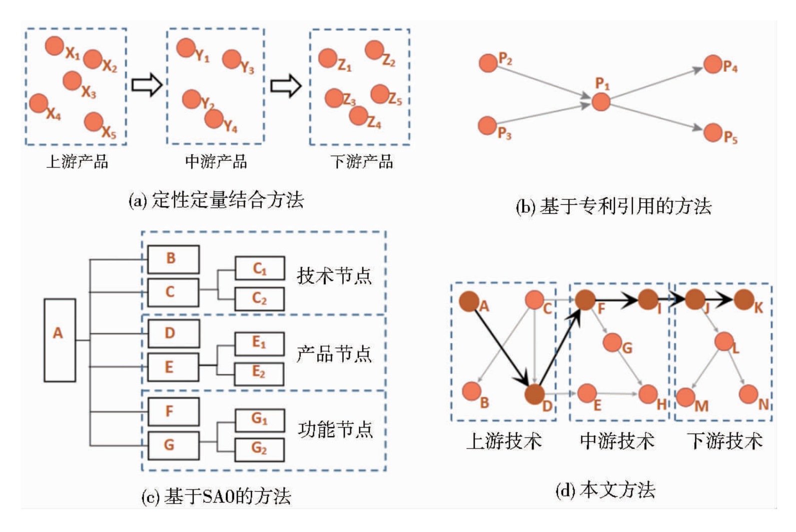


图 1　核心技术链识别方法示意

我们结合图 1 对各种方法进行对比：

2.3.1　定性与定量相结合的方法

图 1 中，(a)表示由该方法识别的核心技术链，其中 X_1-X_5、Y_1-Y_4 和 Z_1-Z_5 分别代表上、中、下游产品相关技术。可以看到，该方法能够从整体上构成产业核心技术链条，但本质上识别的是因产品间的上下游关系而衍生出的技术链接关系[7]，各核心技术环节呈现为围绕产品的技术群或专利群的形式，各核心技术环节间的联系不够紧密和直接。另外，该方法对专家经验的依赖程度较大。

2.3.2　基于专利引用的方法

(b)表示由该方法识别的核心技术链，其中 P1-P5 代表不同的专利，箭头由被引专利指向引用专利。通过具体的部分引文网络可说明这种核心技术链形式的局限。以专利“US2011142242-A1”(P_1)为例，该专利主要描述量子公钥生成技术。在其引证专利和参考专利中，专利“US9736147-B1”(P_4)为“用于在服务器上认证客户端设备的用户的方法”，涉及密钥生成技术的具体应用，属于 P1 的下游技术。而专利“WO2008142816-A1”(P_2)描述通过量子比特空间的预定状态执行量子门矩阵的逆计算来生成量子密钥的技术；专利“JP2002055606-A”(P_3)描述利用代数方程生成公钥的技术；专利“WO2016164325-A1”(P_4)描述通过生成量子弹性随机数生成密钥的技术，它们均属于 P_1 的同类技术。由此可见，基于引文特征识别核心专利链能够有效识别技术领域的核心专利及其之间的知识流动和技术演化，但这种知识流动关系不一定是技术本身的承接关系[7]，而更可能是同类型的技术。因此这种核心技术链难以体现技术间的上下游承接关系。

2.3.3　基于 SAO 的方法

(c)表示基于 SAO 结构的技术树[16]，其中 A-G 为技术要素，C_1-G_2 为各技术要素下的分支，根据 SAO 要素包含的语义信息，可将技术要素划分为技术节点、产品节点、功能节点等多种类型。基于 SAO 构建技术树的方法虽然能够将各技术要素进行具体的类型划分，但难以完整地展现技术间的上下游承接关系并从领域整体层面构建技术链。这是因为，对技术要素的细分会导致要素角色的固化，忽略了技术链上游的输出(如产品)还可以作为技术链下游的输入(如技术)这一普遍的关系，使得技术要素之间的关系难以贯通整个技术领域。

2.3.4　本文方法

(d)表示本文提出的核心技术链识别方法的目标核心技术链。其中 A-N 代表不同的技术要素，箭头表示技术要素间的输入输出关系。图中“A→D→E→F→J→K”表示其中的一条核心技术链，该技术链可以贯穿技术链的上中下游技术(及产品)，能够清晰地展现技术间的承接关系。本文以这种形式作为核心技术链的识别目标。

3　研究设计

3.1　总体思路

针对前述核心技术链识别方法的局限，本文在 SAO 结构和网络分析方法基础上提出了一个基于输入输出型 SAO 网络的核心技术链识别方法，总体思路如下：

3.1.1　设计输入输出型 SAO 结构，描述技术要素间的承接关系

在经典 SAO 结构中，对 S 要素和 O 要素的角色及其之间的关系不做明确规定；在最近提出的多种 SAO 结构中，S、O 角色则固化为产品、技术、属性、效应等角色。这两种 SAO 结构都不利于技术领域上下游技术

要素之间关系的明确和贯通。本文设计了一种新的 SAO 结构，称为“输入输出型 SAO 结构”。输入输出型 SAO 结构将技术要素（S 和 O）概括地区分为“输入要素”和“输出要素”，技术要素间的关系（A）概括为“输入输出关系”，用“输入输出关系”表示技术链中的上下游承接关系。具体地说，本研究根据提取出的 SAO 结构中的 A 的词性（这里的 A 不限于动词，也可以是介词等）和语义，提取 S 与 O 间的包含关系、因果关系、条件关系、指向关系、处理关系等对 S 和 O 进行输入输出方向判别，从而将原有的“S→A→O”关系根据其输入输出关系变为“输入要素→处理要素→输出要素”，即“Input（I）→ Process（P）→ Output（O）”，获得输入输出型 SAO 结构（本文中简称“IPO 结构”），其中 I 和 O 分别为上下游技术要素，P 为技术承接关系。

3.1.2 构建领域知识网络，贯通技术领域上下游的承接关系

在 IPO 结构中，I 和 O 不是固定的角色，一个 IPO 结构的 I 要素可能是上游技术的一个 O 要素。因此，当一个技术领域发展到一定的阶段后，所有专利 IPO 可以构成相互贯通的知识网络。本文借鉴 SAO 网络的构建方法[19]，以 I 和 O 为节点，以 IO 关系在技术领域中出现的频次为重要性，构建有向的、带权重的输入输出型 SAO 网络，本文称为“领域知识网络”，它是核心技术链的“母体”。

3.1.3 化简和分解领域知识网络，获得最重要的技术要素及关系

通过上述方法获得的领域知识网络规模庞大，有可能出现技术要素互为输入输出的相互依赖状态（在网络中体现为“回路”），难以呈现从上游到下游的清晰链条，这使得在专利引文网络中常用于识别核心技术流动的主路径分析法无法应用。为此，本文利用加权 k-Core 方法化简领域知识网络，舍弃其中大多数次要的技术要素和回路，从中提取边权重最大的连通子网络，这个子网络包含了所有专利中出现最频繁的技术要素及其技术关系，本文称为“核心知识子网络”。不过，即使是这个简化后的子网络仍然可能存在回路，难以识别出核心技术链。然而核心知识子网络中的每个回路对应于一个强连通分量，每个强连通分量应该代表其中知识要素的相互依赖关系（非上下游关系）。可通过识别网络的强连通分量将核心知识子网络分解为不含回路的“外部核心知识子网络”与一个或多个含回路的“内部核心知识子网络”，然后根据各强连通分量间的“输入输出”关系确定外部核心知识网络的上下游关系。

3.1.4 利用主路径分析等方法分析核心知识子网络，获得核心技术链

主路径分析是引文网络的一个重要分析工具，它长于分析主流知识的传播和扩散，近期已被用于信息扩散网络中的关键节点识别问题[22]。主路径分析方法的局限在于它不适用于带回路的网络。因此本文通过强连通分量的提取与核心知识子网络的分解，使得外部核心知识子网络不再带有回路（每个回路收缩为一个代表性节点），且整个技术领域的上下游得以贯通；然后，用主路径分析法找出其中最重要的“外部核心技术链”；最后，内部核心知识子网络（回路）可能反映出某些技术环节中具有更为复杂的技术间相互依赖的结构，对每个子网络再次运用上述化简、分解和主路径分析等方法分析其中蕴含的“内部核心技术链”。

总之，本文针对特定产业技术构建基于 IPO 结构的领域知识网络，用加权 k-Core 和主路径分析等方法从领域知识网络中识别和解读该技术的核心技术链。主要技术流程包括数据获取及文本预处理、领域知识网络构建以及核心技术链识别及解读，见图 2。

3.2 具体步骤

3.2.1 数据获取及文本预处理

这一流程分为两个子流程：

（1）数据获取及清洗子流程。本文以 Derwent Innovations Index 专利数据库（简称 DII 数据库）为数据源。根据技术领域的相关知识制定检索表达式，在 DII 数据库中检索并下载相关专利文本，解析专利文献，去除无关信息。

（2）文本预处理子流程。抽取专利文本（本文抽取的是专利摘要），构建领域专利文本集。之后通过指代消歧、分句等步骤进行专利文本的规范化，完成专利文本预处理。

3.2.2 领域知识网络构建

这一流程共分为两个子流程：IPO 结构提取和领域知识网络构建。

（1）IPO 结构提取。这一子流程共分为 3 个步骤：句法分析、提取 IPO 结构三元组、标注 IPO 结构方向及规范化处理。具体步骤如下：

步骤一：句法分析。提取单句中的短语结构树（constituent tree），用于表述句子的成分结构，描述词或词组之间的依存关系。以句子“The quantum computing state detection system comprises a shelving laser to transfer an ytterbium ion from an initial state to a shelved state.”为例，图 3 为该句的短语结构树。



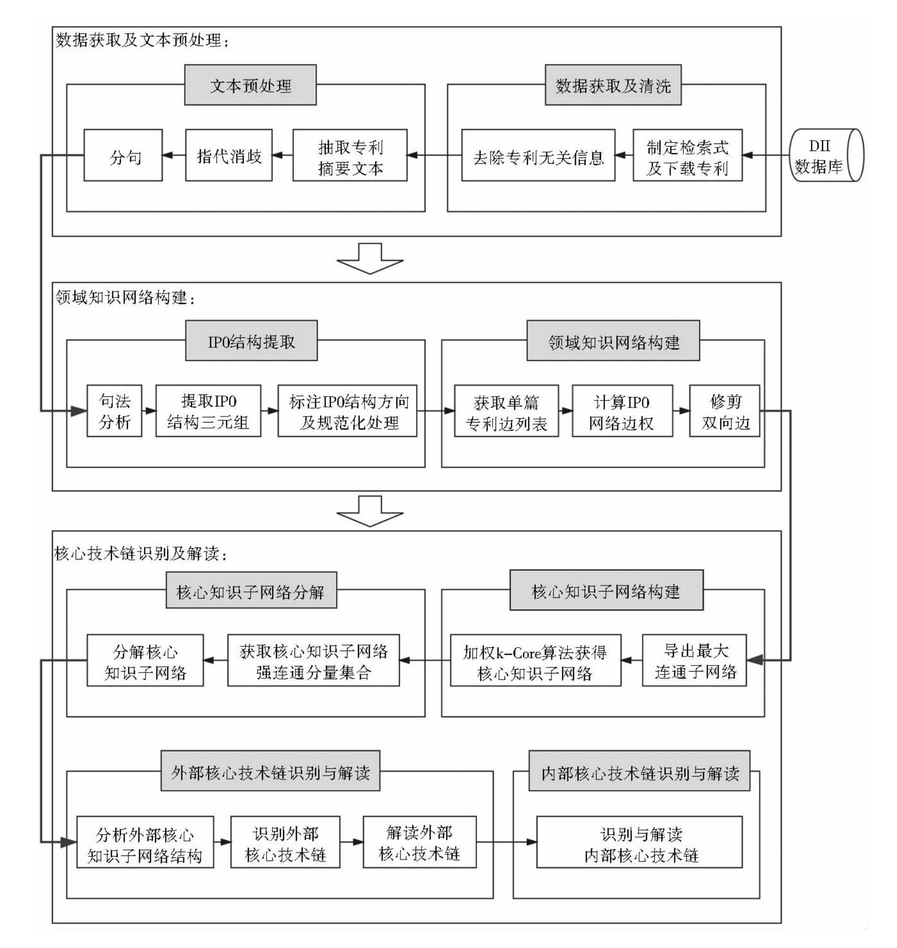


图2 本文核心技术链识别方法的技术路线

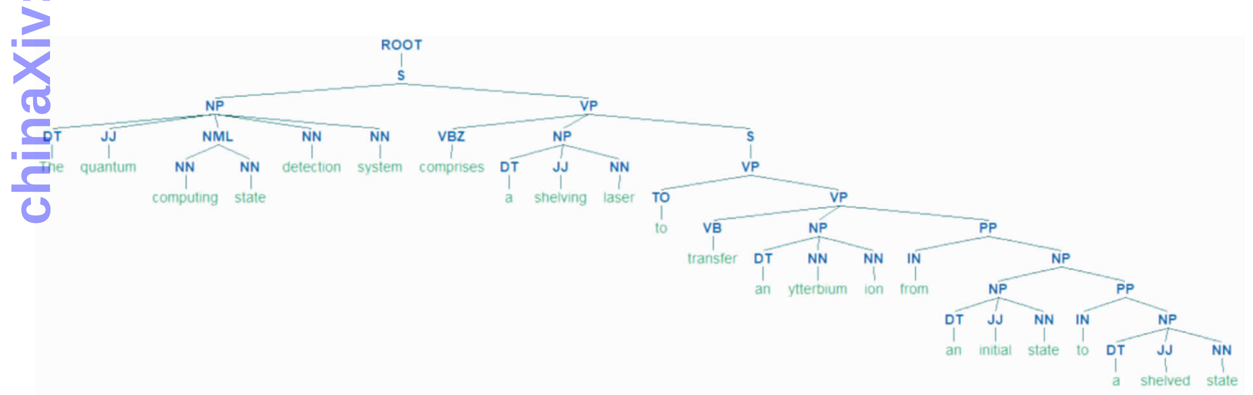


图3 短语结构树示例

步骤二：提取 IPO 结构三元组。首先，根据上一步获得的短语结构树，用正则表达式提取其中的动词、介词和连词作为处理要素（P）（相邻的 P 要素应加以合并）。其次，从短语结构树的根节点出发，根据 P 要素搜寻其最近的父节点及子节点对应的 NP 块，并分别记为 NP1 块，NP2 块，作为输入要素（I）或输出要素（O）。此时获得的[NP_1，P，NP_2]即为 IPO 结构三元组。表1为根据图3提取的 IPO 结构三元组。

表1 IPO 结构三元组示例

NP_1	P	NP_2
The quantum computing state detection system	comprises	a shelving laser
a shelving laser	to transfer	an ytterbium ion
an ytterbium ion	from	an initial state
an initial state	to	a shelved state

步骤三：标注 IPO 结构方向及规范化处理。根据 P 要素的语义构建"IPO 结构方向标注规则表"（详见

表2），利用正则表达式对上一步提取的IPO结构三元组标注方向，三元组中的I要素与O要素的关系多为修饰或限定关系（如P中包含in，of等的关系），故而删去。标注好方向后对IPO结构进行规范化处理，包括词形还原、首字母还原为小写字母以及去除"a""the"等与技术内容无关的无意义的停用词。如上文例句最终的IPO结构如表3所示。

表2 IPO结构方向标注规则

规则	含义	IPO结构方向
P中包含by | based | according to | composed | belong to | depend + | base	该类词常用来描述条件关系，表述为NP_1在NP_2的基础上实现	$NP_2 \rightarrow NP_1$
P中包含from | through | because | due to | out of | via\s	该类词常用来描述因果关系，类似"来自""通过…造成"等，表述NP_1是由于NP_2形成的	
P中包含includ[a-z] + | involv[a-z] + | obtain | \sus[a-z]{1,3} | utiliz[a-z] + | has\) | adopt + | have | having | had | contain | receiv + | compris + | provid | stor + | consist	该类词常用来描述"包含""接收"等包含关系	
P中包含for	该类词常用来描述用途或动作的对象等含义	$NP_1 \rightarrow NP_2$
P中包含to | into | towards	该类词常用来描述到达、指向等表述目的或结果的含义	
P中包含其他能够描述操作、处理等动作的动词，例如get、optimize、predict等	该类词能够明确表述NP1对NP2的操作、处理等动作	

表3 方向标注及规范化处理后的IPO结构示例

I	P	O
shelving laser	comprise	quantum computing state detection system
shelving laser	to transfer	ytterbium ion
initial state	from	ytterbium ion
initial state	to	shelved state

（2）领域知识网络构建。这一子流程以上文得到的I要素和O要素作为知识网络中的节点，P要素作为知识网络的边属性，以I指向O作为边的方向，构建有向的、带权重的领域知识网络。具体步骤如下：

步骤一：获取单篇专利边列表。对于单篇专利，根据上一步IPO提取步骤获得每个单句的IPO结构，从而获得单篇专利的IPO边列表，然后去除单篇专利内重复的输入输出技术关系（即重复的边）。

步骤二：计算IPO网络的边权。汇总技术领域内所有专利的边列表，得到领域边列表，并将I要素与O要素均相同的边合并，将P要素汇总作为该边的属性，将IPO结构所在专利的主入藏号同样作为边的属性，以便于后期结合专利信息对核心技术链进行解读。以IPO网络中各边所包含的P要素的数量衡量边的重要性程度，作为IPO网络的边权。

步骤三：修剪双向边。IPO网络经过加权后可能出现"双向边"，即该条边的两个节点互为输入要素和输出要素，不利于技术链的导出和分析，因此本步骤对这些双向边根据两个节点间输入输出关系的相对重要性进行修剪，去除其中边权较小的方向，将其变换为单向边。如果某一双向边两个方向权重相同，则可以选择同时保留两个方向。经过上述3个步骤，最终完成领域知识网络构建。

3.2.3 核心技术链识别与解读

这一流程分为核心知识子网络构建、核心知识子网络分解、外部核心技术链识别与解读以及内部核心技术链识别与解读4个子流程。

（1）核心知识子网络构建。这一子流程的目标是简化领域知识网络，从中提取技术领域的核心技术要素和技术关系。具体步骤如下：

步骤一：导出其中的最大连通子网络，去除其他重要性较低的小的连通子网络，作为核心技术链的主要知识要素。

步骤二：通过加权k-Core算法获得核心知识子网络。k-Core算法是一类用于在网络中找出符合指定核心度的紧密关联的子网络结构的方法。在经典k-Core中，每个顶点至少具有k的度值，且所有顶点都至少与该子网络中的k个其他节点相连，强调节点的重要性。而核心技术链是产业链中增值最大且不可或缺的技术环节，从世界范围看，核心技术链所涉及的技术必然是产业研发竞争最为激烈的，专利也通常最为集中。这些专利所揭示的共同技术联系越多，在知识网络中的权重也就越大，越有可能成为核心技术链的边。因此本文使用更强调边的重要性的加权k-Core算法，即通过设定最大边权阈值e，使得k-Core中的每个顶点与邻接节点的边权至少有一条大于e，从而得到保留了最重要技术关系的领域核心知识子网络。

（2）核心知识子网络分解。这一子流程的目标是将核心知识子网络分解为外部和内部两类核心技术子网络，为核心技术链识别与解读做准备。具体步骤为：

步骤一：获取核心知识子网络中的强连通分量集合。导出其中的强连通分量，获取核心知识子网络中

的强连通分量{T},以便后面对各强连通分量的技术内容进行具体识别和解读。假定某领域的核心知识子网络如图4所示。由于B、C、D和E这4个节点之间存在强连通关系,可以得到该网络中的强连通分量集合{A,{B,C,D,E},F,G}。

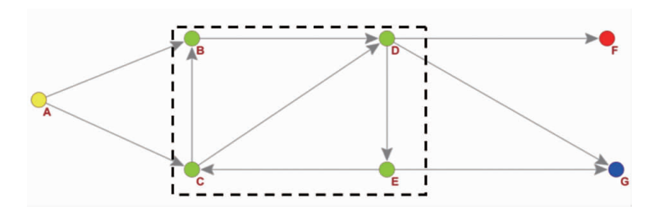


图4　核心知识子网络示例

步骤二:分解核心知识子网络。将每个强连通分量收缩为一个节点,选取其中一个代表性节点作为标签,收缩后的领域核心知识子网络称为"外部核心知识子网络",用以描述技术领域整体层面上的技术知识。而那些具有多个节点的强连通分量则被称为"内部核心知识子网络",用于具体描述相互依赖的内部技术流程。例如,图4中的强连通分量{B,C,D,E}收缩为一个节点(设用#B标记),与A,F和G一起构成外部核心知识子网络(见图5),而{B,C,D,E}及其连边则构成内部核心知识子网络。

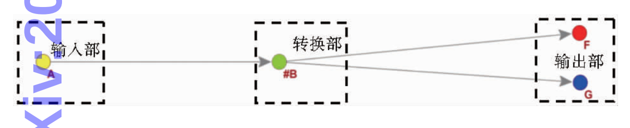


图5　外部核心知识子网络示例

(3)外部核心技术链识别与解读。这一子流程分析外部核心知识子网络的结构,识别外部核心技术链,然后解读每个外部核心技术链的具体内容。分为以下3步:

步骤一:分析外部核心知识子网络的结构。由于去除了强连通分量,外部核心知识子网络的结构变得更加清晰,网络一般呈现一个或多个链状或树形结构。如图5所示,外部核心知识子网络可以根据各节点的输入输出情况划分为输入部、转换部、输出部3个部分(也可能更多)。输入部包含该子网络的源点集合,主要描述技术领域的原理与部件;转换部通常由若干强连通分量构成,处于技术领域中间或核心位置,主要描述内部核心技术的操作与流程;输出部包含该子网络的汇点集合,主要描述技术领域的功能与结果。

步骤二:识别外部核心技术链。根据外部核心知识子网络的结构,分析输入部与输出部原理与部件、操作与流程、功能与结果的关系,运用主路径分析方法识别外部核心技术链。首先,运用主路径分析法求得外部核心知识子网络的全局主路径[23],作为一条核心技术链。该技术链在所有外部核心技术链中总边权最高,因此所描述的技术关系较为重要,描述了技术领域中最主流的技术流程,因此我们称其为"主流外部核心技术链"。其次,利用关键主路径分析方法[23],将外部核心知识子网络中边权重较高的若干边作为关键边,搜索和识别其他主要的外部核心技术链,我们称其为"主要外部核心技术链"。

步骤三:解读外部核心技术链。结合各外部核心技术链中重要边的来源专利对主流外部核心技术链和主要外部核心技术链进行解读。首先找到各外部核心技术链中权重较大的边所在的IPO结构,根据IPO结构中的P要素确定I和O间的对应关系,然后根据IPO结构的来源专利的摘要原文,对各个外部核心技术链进行深入解读。

(4)内部核心技术链识别与解读。这一子流程将前述分析方法嵌套式地运用于内部核心知识子网络S,以识别和解读内部核心技术链。包含4个步骤:

步骤一:参照3.2.3小节的(1)部分,提高e值,提取S的加权k-Core,得到更为核心的技术子网络,记为S′。

步骤二:参照3.2.3小节的(2)部分,获取S′的强连通分量集合,并对S′进行收缩。

步骤三:参照3.2.3小节的(3)部分,通过主路径分析方法对S′的"外部"核心技术链进行识别,并依据各核心技术链来源专利对它们进行解读。

步骤四:对S′的强连通分量集合进一步分析与解读,识别并解读其中重要的"内部"核心技术链。

4　案例研究

量子计算是一种遵循量子力学规律进行高速运算、存储、处理信息的新型计算,其储存能力强,运算速度快,将带来现有计算能力质的飞跃[24]。我国也高度重视量子计算技术的发展,并将其列入国家发展规划中。而今,量子计算领域研发进展迅速,越来越多的国家、企业和研发机构关注量子计算技术的研究。目前量子计算仍处于技术验证和原理样机研制的关键阶段[25]。因此,识别量子计算领域的核心技术链可以帮助技术管理部门构建量子计算领域核心技术架构,帮助科研机构寻找技术薄弱点及确定技术攻关方向具有实际意义。

4.1　数据获取及文本预处理

本文选择专利数据,以DII数据库为数据来源,识

别量子计算领域的核心技术链。德温特手工代码（DMC）是德温特索引专家根据专利摘要和全文给出的标注，用以揭示和表达技术类别，相比国际专利分类号（IPC）分类更为明确。因此本文将德温特手工代码（DMC）作为检索字段，通过查阅相关文献，确定量子计算的德温特手工代码为 T01-E05Q，因此制定检索表达式为“MAN = T01-E05Q”，数据检索时间为 2021 年 1 月 2 日，共获得量子计算领域 1998 年以来的 2 303 项专利。图 6 展示了该领域各年的专利申请数量变化情况。根据对量子计算领域各年专利申请数量变化的统计，该领域专利申请数量自 2002 年开始呈现出递增趋势，2015 年之前增长趋势较为缓慢，2015 年以后呈现迅猛发展，这表明了量子计算领域的技术越来越受到大家的重视。通过对专利申请国家或地区的统计，在中国申请的专利数量仅次于美国，居第二位，这说明中国在量子计算领域的研发力量和市场前景较好。

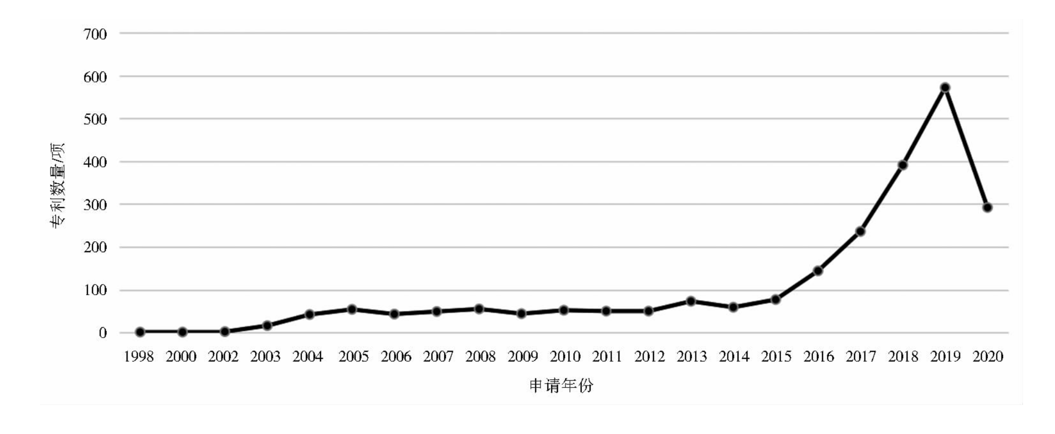


图 6　量子计算领域专利年度变化趋势

之后运用 Python 及其自然语言工具包编写代码抽取专利摘要文本，去除无关专利信息，清洗数据。而后通过分句、指代消歧等步骤实现文本预处理，获取领域专利摘要文本集。

4.2　领域知识网络构建

在获得量子计算领域的专利摘要文本集后，按照前述 IPO 结构提取步骤，借助 Standford NLP Parser 工具、NLTK 工具包以及正则表达式提取该技术领域的 IPO 结构。

之后，经过获取单篇专利边列表、计算 IPO 网络边权、修剪双向边 3 个步骤构建量子计算领域知识网络。该网络共包括 22 782 个节点、29 504 条边。

4.3　核心技术链识别与解读

4.3.1　核心知识子网络构建与分解

利用复杂网络分析软件（如 Pajek）对量子计算领域 IPO 知识网络进行分析，可知其包含弱连通分量共 1 954 个，除最大连通子网络外各弱连通分量较小，包含的知识要素较少，因此选择其中的最大连通子网络识别核心技术链。该最大连通子网络共包含 18 080 个节点，26 490 条边。之后提取该网络的加权 k-Core（边权阈值 e 设为 5），得到量子计算领域的核心知识子网络（见图 7），它共有 69 个节点、189 条边，每个节点至少有一条权重大于 5 的邻接边。

其次，对核心知识子网络进行强连通性分析，得到该网络的 37 个强连通分量集合（见图 7）。其中 36 个分量都是单个节点，只有虚线框中的 33 个节点形成最大的强连通分量，它们包含的技术从输入输出角度看互相依赖，都是有关量子计算装置原理和构造的，因此将“#device”作为该强连通分量（内部核心知识子网络）的标签，获得收缩后的外部核心知识子网络如图 8 所示（图 8 边上的数值代表边权）。

4.3.2　外部核心技术链识别与解读

图 8 中外部核心知识子网络可以根据各节点的输入输出情况划分为输入部、转换部、输出部 3 个部分。通过主路径分析得到包含 IPO 关系的外部核心技术链，再结合各链所涉及的专利对其进行详细解读。

首先，通过全局主路径分析可以获得一条“主流外部核心技术链”（见图 9 中标注加粗部分）。之后利用关键主路径分析找到所有包含权重排名前 15 的边的关键主路径。由于这些主路径有很多重合边，为方便起见，我们将它们合并起来形成“主要外部核心技术链”，见图 9。

外部核心技术链的解读分为两部分。首先，对主流外部核心技术链进行解读，该技术链包含的 IPO 结构以及其来源专利见表 4。



图 7　量子计算领域的核心知识子网络

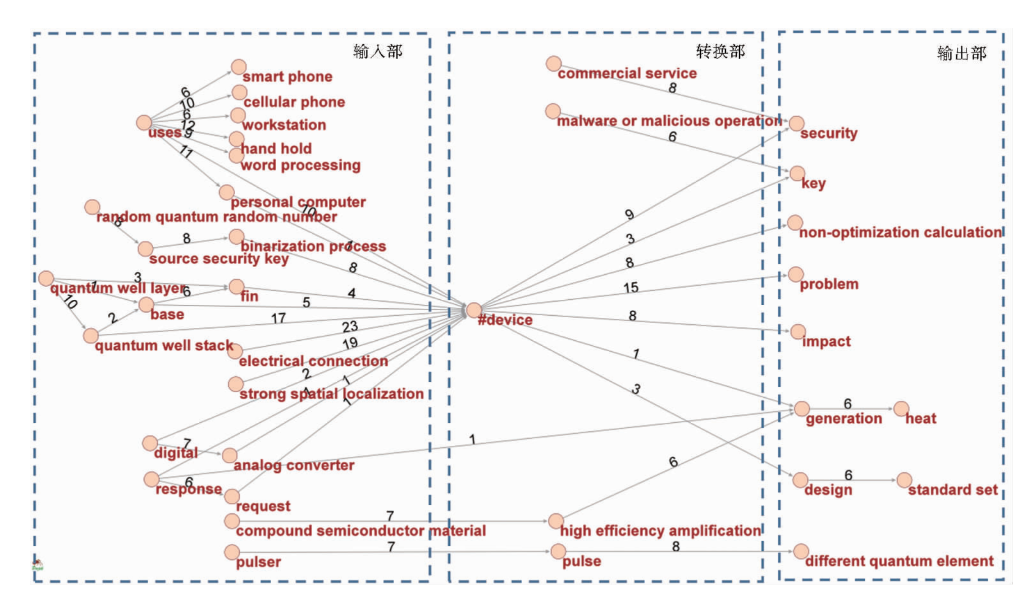


图 8　外部核心知识子网络

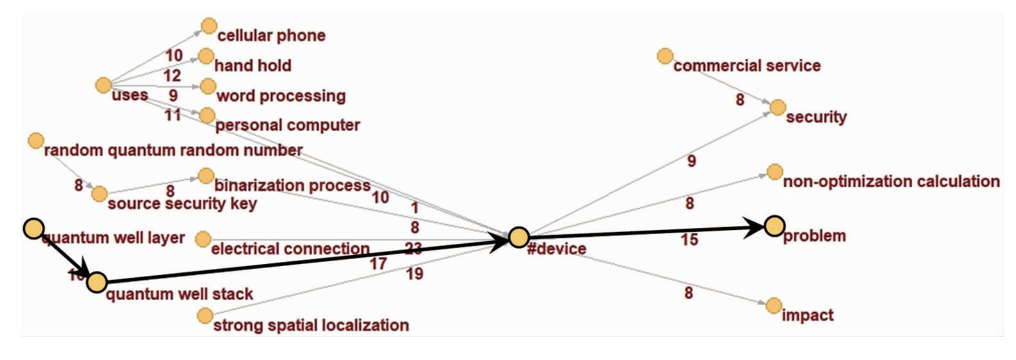


图 9　主要外部核心技术链(含主流外部核心技术链)

表 4　主流外部核心技术链所含 IPO 结构以及其来源专利

边	处理要素(P)	来源专利(DII 主入藏号)
quantum well layer→quantum well stack	include;which include;that include;comprise	201785401J;201785401G;201826641K;201825701S;201813310S;201785401H;2019540027;201933093W;2019137543;201913755F
quantum well stack→#device	include; and include; have; comprise;which include	201785401J;2019699502;201785401G;201826641K;201913755C;201913755F;2018012832;201900097P;201826641L;201825701S;201825701Y;201813310S;2017854019;201785401F;202096994Y;202096994K;2019540027;2019540029;2018257022;2019A73587;2019137543
#device→problem	to;complete;to solve;represent by	2008D21626; 2011A19660; 201994505K; 2020B6443G; 2020923016; 202086297X; 202068913W; 201894731P; 2014N52167; 201952045R; 201992268W; 2020B83555; 2020B83555;201894731P;2019A3420N

依据该表,可以对主流外部核心技术链进行如下分析和解读:

(1)边“quantum well layer → quantum well stack”描述量子阱层与量子阱堆叠之间的关系,结合该边的处理要素以及所在原专利,可知该边主要描述量子阱堆叠的物理结构,即量子阱堆叠由量子阱层构成。

(2)边“quantum well stack → #device”描述量子阱堆与量子计算设备之间的关系,结合该边的处理要素以及所在原专利,可知该边主要表述量子点器件的物理结构,量子点器件由量子阱堆叠、掺杂层和阻挡层构成,因此量子阱堆叠是量子点器件的构成部分。

(3)边“#device → problem”描述量子计算设备与问题之间的关系,两个要素所在 IPO 结构中的处理要素包括“to”“complete”“to solve”等表述“完成”“解决”等含义的词,结合该边所在原专利,可知该边主要表述量子计算设备能够解决的问题,包括量子计算设备物理结构优化问题以及基于量子计算解决的整数规划、组合优化、大数据搜索优化以及多元二次问题等数学优化问题。

综上所述,主流外部核心技术链的上游是量子点器件的物理结构,包含量子阱堆叠,其中量子阱堆叠由量子阱层构成;技术链下游是量子计算设备可解决的问题,主要包括量子计算设备物理结构优化问题和各种复杂的数学最优化问题。这是迄今为止量子计算最主流的技术实现方案和应用领域[24]。

其次,参照主流外部核心技术链的解读步骤对这些主要外部核心技术链一并解读。从图 9 可以看出,几乎所有主路径都从外部核心技术子网络的输入部出发,途经“#device”这一转换部,抵达输出部。由于篇幅所限,这里将主要外部核心技术链的技术内容概括为以下两方面:

(1)输入部的子技术链主要描述量子计算设备的物理结构或工作原理等。其中,子链“random quantum random number → source security key → binarization process → #device”描述基于量子计算设备提高网络安全性的原理;“electrical connection → #device”描述量子点设备的集成方式;“quantum well layer → quantum well stack → #device”(在全局主路径上)描述量子点设备的物理组成结构;“strong spatial localization → #device”描述量子点设备的空间特性。

(2)输出部的子技术链主要描述量子计算设备的用途,包括在量子计算设备内部优化上的用途以及其能够实现的功能。其中,子链“#device → security”描述其在提高网络安全性上的功能;“#device → non-optimization calculation”描述其在解决非最优化计算问题方面的功能;“#device → problem”(在全局主路径上)描述其可以解决的量子计算设备内部物理结构优化问题和一系列数学优化问题;“#device → impact”描述其在量子计算设备内部优化上的用途,如通过。

需要说明的是,通过上述主路径分析有可能得到一些看似并非核心技术链的主路径。例如图 9 中从“uses”节点出发分别到“cellular phone”“hand hold”“word processing”和“personal computer”的边主要描述处理器等设备在个人电脑、便携式电话等方面的用途,而“uses → #device”主要描述这些设备在量子计算设备中的功能。前者不涉及量子计算的核心技术,而后者与量子计算设备中的物理组成结构等有关。

4.3.3　内部核心技术链识别与解读

识别和解读量子计算领域外部核心技术链后,接下来对内部核心知识子网络(“#device”分量)加以分析,识别和解读量子计算领域的内部核心技术链。首先对“#device”分量再次提取加权 k-Core(边权阈值 e 提高到 9),可以获得“#device”分量更为核心的知识子网络。然后,按照主路径分析方法对其进行强连通分析,收缩其强连通分量,得到 “#device”分量的“外部”核心知识子网络 S′ 以及“#processor”和“#quantum processing device”两个“内部”核心知识子网络,分别记为 S1″和 S2″,如图 10 所示:

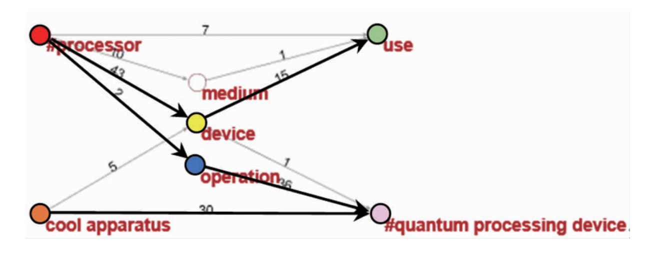


图 10 “#device”分量的“外部”核心知识子网络 S′

再次，利用关键主路径分析方法识别 S′中包含权重排名前 3 的边的主路径，共包含 3 条“主要内部核心技术链”（图 10 中标注加粗部分）。

根据这 3 条主要内部核心技术链所在原专利对其进行以下解读：

（1）核心技术链“#processor → device → use”上游描述量子计算设备中的量子点设备或量子控制设备的物理结构，下游描述该量子点设备或量子控制设备在量子计算设备中的功能。

（2）核心技术链“#processor → operation → #quantum processing device”上游描述量子处理器能够执行的操作，下游描述该操作产生的影响，包括量子退相干以及散射效应等。

（3）核心技术链“cool apparatus → #quantum processing device”主要描述冷却设备在量子计算设备中的功能，包括将量子处理器保持在较低温度以避免量子位退相干和减少散射效应。

最后，对“#processor”和“#quantum processing device”两个“内部”核心知识子网络 S1″和 S2″分别识别其中包含的内部核心技术链，并进行如下解读：

（1）S1″（“#processor”分量）中主要的内部核心技术链为“qubit decoherence → effect”，主要描述冷却设备在量子技术计算设备内部起到的避免散射效应的功能。

（2）S2″（“#quantum processing device”分量）主要的内部核心技术链为“processor →computer executable component → memory → system”，该技术链上游描述量子处理器对计算机可执行组件的执行技术，中游描述计算机可执行组件将指令放入存储器，下游指出存储器是量子计算系统的构成部分。

4.3.4 结果验证

综合外部核心技术链和内部核心技术链的识别结果，本文总结了量子计算领域的核心技术链（见图 11）。可以看出量子计算领域的核心技术链分上、中、下游 3 部分，其中上游主要描述量子计算设备的相关原理和物理结构，中游主要描述量子计算设备的内部技术，下游主要描述量子计算设备的应用和功能。

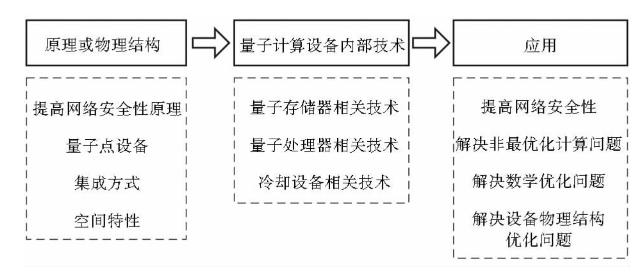


图 11 量子计算领域核心技术链总结

为了进行结果验证，项目组向两位量子计算领域的教授展示了运用该方法的识别结果（主要是图 9－图 11），他们都认为识别出的核心技术链较为完整，能够准确反映现阶段量子计算产业发展的主要技术及其关系。其中一名教授指出量子计算机的某些技术路线（如超导、离子阱和光子等）没有在本文识别的核心技术链中得到体现。在与该教授的讨论中，发现这些技术路线在高水平论文中出现较多，但在专利中体现较少，而且多条技术路线尚未收敛为主流技术，因此没有出现在当前的核心技术链中。此外，笔者还查阅了一些总结量子计算领域技术发展状况的文献，通过研究结果的对比来评价本文提出的核心技术链识别方法的可行性和有效性。

（1）量子计算领域上游核心技术链描述量子计算设备的原理或物理结构，包括提高网络安全性的原理和量子点设备、量子设备间的集成方式及其空间特点等。其中量子计算设备提高网络安全性的原理，主要涉及“量子密钥分发技术”，面向非对称密码系统，根据相关研究，该技术被认为可以提高量子通信的保密性，属于量子计算中的关键技术[26-27]；量子点设备被认为是能够实现量子计算的一种方式[26,28-30]，而量子点材料及器件也是目前国际上最前沿的研究课题之一，其物理构成中包含的量子阱材料是新一代固态量子器件的基础材料[31]；另外集成方式[29]和空间特性[32]也被认为是当前量子计算领域研究的关键。

（2）量子计算领域核心技术链的中游主要介绍量子计算设备的内部技术，包括量子存储器和量子处理器（量子芯片）的数据处理技术以及冷却设备相关技术等。量子处理器的物理实现以及数据处理技术均是量子计算的关键技术[24-25,31]。当前固态量子芯片也面临着退相干严重等问题，这也是实现量子计算机的困难之处[30,33]。降低量子计算设备的温度，保持长时间的量子态相干是当前最大的技术瓶颈[28,33]，因此冷却

设备相关技术也是如今研究的重点。

（3）量子计算领域核心技术链的下游主要介绍量子计算设备的应用。根据相关研究，量子计算的应用主要集中在大规模数据处理[33]、信息安全保护[28,33]、最优化问题解决[26,33]等方面，这与本文的分析结果是吻合的。

综上所述，本文识别出的量子计算领域核心技术链与相关研究的结果高度一致。

5 结论

识别产业中的核心技术链对于打造完整的技术链和产业链有重要的理论和现实意义。然而，现有方法在自动和完整地识别核心技术链方面尚显不足。本文提出一种基于输入输出型 SAO 网络的核心技术链识别方法，以从专利文本中提取的输入输出型 SAO 结构为技术要素，构建技术领域知识网络，通过加权 k-Core 算法获取包含重要技术关系的核心知识子网络；然后运用强连通分量提取和主路径分析等方法识别出其中蕴含的外部和内部核心技术链；最后，以量子计算领域为例，对该领域的核心技术链进行识别，并依据专利原文对核心技术链进行解读。通过与量子计算领域的专业研究成果加以对比，证实了本文所提出方法的有效性。

本文提出的核心技术链识别方法的主要优点为：

（1）输入输出型 SAO 结构相比经典的 SAO 结构更能明确体现技术间的承接关系，较适合技术链的识别和分析；输入输出型 SAO 结构提取的自动化程度高，能显著减少专家知识的介入。

（2）领域知识网络以技术关系在领域中出现频次作为边权，能够按照重要性程度获得主要和次要等不同等级的核心技术链，具有较强的灵活性。

（3）通过加权 k-Core 算法和强连通分量的预处理，将主路径分析方法的应用领域加以扩展，使其能够应用于含回路的知识网络。

本文尚存在一些不足之处，一是专利摘要不能够完全把握技术细节；二是目前技术要素间的输入输出关系提取规则还存在一定的模糊性和多重性；三是仅靠专利数据来识别核心技术链发现的是相对比较成熟的当前核心技术，而在科学论文等数据源中可能蕴含更为前沿的核心技术链。在未来的研究中，除了通过引入异构数据扩大技术数据源以及完善输入输出型 SAO 结构的提取规则等途径提高核心技术链的识别准确度之外，还应对识别出的核心技术链进行更加深入的分析，如竞争力比较和技术链动态演化分析等，使核心技术链研究更好地为提升我国产业技术水平服务。

参考文献：

[1] 吴金明，邵昶. 产业链形成机制研究——“4+4+4”模型［J］. 中国工业经济，2006(4)：36-43.
[2] 赵君丽，钟杰，王立新. “三链”视角下战略性新兴产业集群发展策略研究——以西宁碳纤维产业集群为例［J］. 青海社会科学，2020(5)：96-105.
[3] 洪勇，苏敬勤. 发展中国家核心技术产业链与核心技术链协同发展研究［J］. 中国工业经济，2007(6)：38-45.
[4] 陈思，赵宇翔，朱庆华. 基于技术链的产业技术竞争情报服务模式探析［J］. 情报理论与实践，2020，43(5)：31-37.
[5] 远德玉. 产业技术界说［J］. 东北大学学报(社会科学版)，2000(1)：22-25.
[6] 柳琦，丁云龙. 产业集群的技术成因分析［J］. 中国科技论坛，2005(4)：122-125.
[7] 高汝熹，纪云涛，陈志洪. 技术链与产业选择的系统分析［J］. 研究与发展管理，2006(6)：95-101.
[8] 黄立业，赵辉，王坚，等. 基于专利分析的产业竞争情报分析框架研究［J］. 情报科学，2015，33(4)：59-63.
[9] 朱瑞博. 核心技术链、核心产业链及其区域产业跃迁式升级路径［J］. 经济管理，2011，33(4)：43-53.
[10] 方思，李国秋. 全球无人驾驶汽车专利分析——从产业链和技术链的二维角度［J］. 竞争情报，2016，12(5)：27-36.
[11] 吴菲菲，栾静静，黄鲁成，等. 产业技术链视角下我国动力电池产业问题分析与对策研究［J］. 科技进步与对策，2017，34(2)：58-63.
[12] 袁晓东，鲍业文. “中兴事件”对我国产业发展的启示：基于专利分析［J］. 情报杂志，2019，38(1)：23-29.
[13] WEI Y M，KANG J N，YU B Y，et al. A dynamic forward-citation full path model for technology monitoring：an empirical study from shale gas industry［J］. Applied energy，2017，205：769-780.
[14] 赵蓉英，李新来，李丹阳. 专利引证视角下的核心专利研究——以人工智能领域为例［J］. 情报理论与实践，2019，42(3)：78-84.
[15] 戚筠，唐恒，石俊国. 基于小世界网络特性的核心技术识别研究——以石墨烯为例［J］. 情报杂志，2020，39(2)：50-55.
[16] CHOI S，PARK H，KANG D，et al. An SAO-based text mining approach to building a technology tree for technology planning［J］. Expert systems with applications，2012，39(13)：11443-11455.
[17] 翟东升，夏军，张杰，等. 基于专利特征抽取的技术树构建方法研究［J］. 情报学报，2015，34(7)：717-724.
[18] 郭俊芳，汪雪锋，李乾瑞，等. 一种新型的技术形态识别方法－基于 SAO 语义挖掘方法［J］. 科学学研究，2016，34(1)：13-21.
[19] YANG C，HUANG C，SU J. An improved SAO network-based method for technology trend analysis：a case study of graphene［J］. Journal of informetrics，2018，12(1)：271-286.

[20] 赖朝安,侯延行,文雄辉. 基于专利 SAO 链的技术预见方法研究——以工业机器人领域为例[J]. 科技管理研究,2020,40(7):171－179.

[21] 李晓曼,张学福,宋红燕,等. 专利文献技术要素识别方法研究——以纳米肥料领域为例[J]. 图书情报工作,2020,64(6):59－68.

[22] 魏莹,李峰. 基于主路径分析的关键节点识别与谣言扩散抑制研究[J]. 情报科学,2018,36(6):13－19.

[23] LIU J S,LU L Y Y. An integrated approach for main path analysis: development of the hirsch index as an example[J]. Journal of the American Society for Information Science and Technology,2012,63(3):528－542.

[24] 王立娜,唐川,田倩飞,等. 全球量子计算发展态势分析[J]. 世界科技研究与发展,2019,41(6):569－584.

[25] 张海懿,崔潇,吴冰冰. 量子计算技术产业发展现状与应用分析[J]. 信息通信技术与政策,2020(7):20－26.

[26] 郑建国,覃朝勇. 量子计算进展与展望[J]. 计算机应用研究,2008(3):641－645.

[27] ZHANG J,TIAN Q,TANG C,et al. Study on worldwide development and trends of quantum technologies based on patent data[J]. International journal of information and education technology,2020,10(3):239－244.

[28] 钟诚,陈国良. 量子计算及其应用[J]. 广西大学学报(自然科学版),2002(1):83－86.

[29] 陈平形,吴伟,吴春旺,等. 量子计算的研究现状和发展动向[J]. 国防科技,2014,35(6):3－6.

[30] 田倩飞,王立娜,唐川,等. 基于文献计量的量子计算研究国际发展态势分析[J]. 科学观察,2019,14(6):1－9.

[31] 刘小平,李泽霞. 基于共词分析的量子信息学前沿热点分析[J]. 科学观察,2014,9(5):13－22.

[32] ZHONG J,OLEKHNO N A,KE Y,et al. Photon-mediated localization in two-level qubit arrays[J]. Physical review letters,2020,124(9):1－6.

[33] 李联宁. 穿越未来世界——量子计算机[M]. 北京:清华大学出版社,2019.

作者贡献说明:

任海英:制定选题,设计研究思路,确定研究方法,修改论文并定稿;

李真:研究相关方法及理论,收集数据,确定相关方法,撰写论文初稿。

Research on Identification Method of Core Technology Chain Based on an Input-Output SAO Network: A Case of Quantum Computing Technologies

Ren Haiying Li Zhen

School of Economics and Management, Beijing University of Technology, Beijing 100124

Abstract: [Purpose/significance] This paper identifies the core technology chain of a certain technology field based on patent data, which is important for clarifying the structure of industrial core technology, analyzing the weak links of China's core industry chains, determining the direction of technology research, and improving the technology chains and industry chains. [Method/process] This method improved the classic SAO, and proposed a core technology chain identification method based on an input-output SAO network. It extracted input-output SAO structures from the domain patent data as technical elements, and constructed a domain knowledge network according to the input-output relationship between the technical elements. It used the weighted k-Core method to obtain the core knowledge subnetwork which contained the main technical relations. It applied main path analyses to identify the external core technology chain and the internal core technology chain in the core knowledge subnetwork, and interpreted the core technology chain with reference to the patents involved. [Result/conclusion] This paper applied the proposed method in the field of quantum computing to identify the external and internal core technology chains. The accuracy of the results is verified by expert evaluation and comparison with the research of relevant scholars. The proposed method enables the identification of coherent technology chains, and has high degree of automation and flexibility.

Keywords: core technology chain SAO structure knowledge network main path analysis weighted k-Core